

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-296522
 (43)Date of publication of application : 29.10.1999

(51)Int.Cl. G06F 17/30

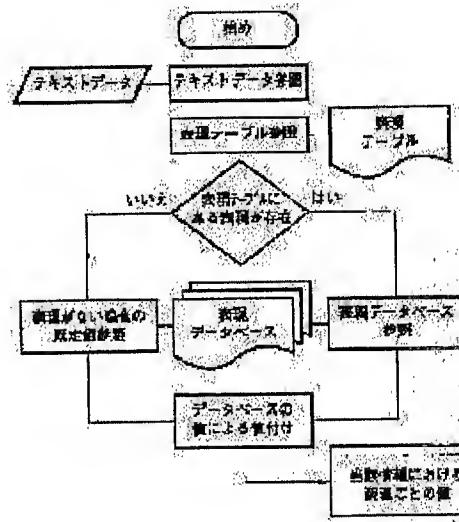
(21)Application number : 10-102886 (71)Applicant : NIPPON TELEGR & TELEPH CORP
 <NTT>
 (22)Date of filing : 14.04.1998 (72)Inventor : KAKUTA HIROKAZU
 SUMI RYUICHI

(54) SYSTEM AND METHOD FOR DECIDING ATTRIBUTE OF INFORMATION AND RECORD MEDIUM WHERE ATTRIBUTE DECIDING PROGRAM IS RECORDED

(57)Abstract:

PROBLEM TO BE SOLVED: To provided the system and method for information attribute decision making which decide the attribute of information more precisely and faster and the record medium where the attribute deciding program for information is recorded.

SOLUTION: A word representing the attribute of information included in information consisting of a text or elements which can be converted into the text is extracted from the information and this extracted work is given a numeral according to specific algorithm; and obtained numerals are combined according to other specific algorithm to obtain a different numeral, with which the attribute of the information is quantitatively decided.



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-296522

(43)公開日 平成11年(1999)10月29日

(51)Int.Cl.⁶
G 0 6 F 17/30

識別記号

F I
G 0 6 F 15/403

3 4 0 A

審査請求 未請求 請求項の数3 O L (全7頁)

(21)出願番号 特願平10-102886

(22)出願日 平成10年(1998)4月14日

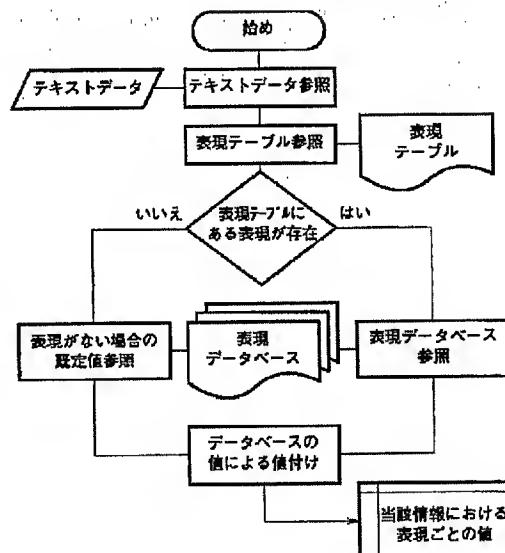
(71)出願人 000004226
日本電信電話株式会社
東京都千代田区大手町二丁目3番1号
(72)発明者 加来田 裕和
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内
(72)発明者 角 隆一
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内
(74)代理人 弁理士 鈴江 武彦 (外2名)

(54)【発明の名称】 情報の属性判定システム及び方法並びに情報の属性判定プログラムを記録した記録媒体

(57)【要約】

【課題】本発明の課題は、より精度良く高速に情報の属性を判定する情報の属性判定システム及び方法並びに情報の属性判定プログラムを記録した記録媒体を提供することにある。

【解決手段】本発明は、テキストもしくはテキストに変換可能な要素から構成される情報に含まれる情報の属性を表わす言葉を当該情報から抽出し、この抽出された言葉に所定のアルゴリズムにしたがって数値を付け、さらに、得られた数値を他の所定のアルゴリズムにしたがって組み合わせ別の数値を得、この得られた数値によって当該情報の属性を定量的に判定することを特徴とする。



【特許請求の範囲】

【請求項1】 テキストもしくはテキストに変換可能な要素から構成される情報に含まれる情報の属性を表わす言葉を当該情報から抽出する言葉抽出手段と、この言葉抽出手段で抽出された言葉に所定のアルゴリズムにしたがって数値を付ける第1の数値付与手段と、この第1の数値付与手段で得られた数値を他の所定のアルゴリズムにしたがって組み合わせ別の数値を得る第2の数値付与手段と、この第2の数値付与手段で得られた数値によって当該情報の属性を定量的に判定する判定手段とを具備することを特徴とする情報の属性判定システム。

【請求項2】 テキストもしくはテキストに変換可能な要素から構成される情報に含まれる情報の属性を表わす言葉を当該情報から抽出するテキストデータ参照ステップと、このテキストデータ参照ステップで抽出された言葉を表現テーブルと参照し、表現テーブルにある表現が存在するか否かを判別する表現テーブル参照ステップと、この表現テーブル参照ステップで判別された言葉を表現データベースと参照し、表現データベース値による値付けをする表現データベース参照ステップと、この表現データベース参照ステップで得られた数値を計算式データベースと参照し、表現ごとの値を計算式へ代入して当該情報の値の計算をする計算式参照ステップと、この計算式参照ステップで得られた数値によって当該情報の属性を定量的に判定する判定ステップとを具備することを特徴とする情報の属性判定方法。

【請求項3】 コンピュータに、テキストもしくはテキストに変換可能な要素から構成される情報に含まれる情報の属性を表わす言葉を当該情報から抽出する第1の手順、この第1の手順で抽出された言葉に表現ごとの値付けをする第2の手順、この第2の手順で得られた表現ごとの値を計算式へ代入して当該情報の値の計算をする第3の手順、この第3の手順で得られた数値によって当該情報の属性を定量的に判定する第4の手順、を実行させるための情報の属性判定プログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、テキストまたはテキストに変換可能な要素で構成され、蓄積または送受される情報を対象として、当該情報の属性（たとえば重要性や緊急性など）を定量的に表わすことによって情報の属性を容易に判定できるようにして、情報の利用者が効率よく情報を処理することを可能とする情報の属性判定システム及び方法並びに情報の属性判定プログラムを記録した記録媒体に関するものである。

【0002】

【従来の技術】従来、情報の利用者が効率よく情報を処理できるように、例えばWWWや電子メールやNEWS、メーリングリスト、電話、FAX、テレビなどの情報メディア、および例えば情報検索（ディレクトリサービス、ページ検索サービスなど）や情報配信サービス（電子メール新聞）などのサービスでは、利用者自ら嗜好に合せて情報を選択・並べ替えを行ったり、利用者が年齢、性別やキーワード（「DOS/V」、「イタリア料理」など対象領域を表わす単語）を登録することによって選択・並べ替えを行ったりする方法がある。たとえば、テレビのチャンネルをあわせるという方法や、キーワードを登録してそのキーワードに合った情報をFAXで送信するという方法である。また、最近の利用者の傾向（つまり流行）から提示順位を付けたり、利用者の特徴を示す項目が似通った他の人の傾向から、その利用者への情報の提示順序を自動で判定したりする方法がある。

【0003】

【発明が解決しようとする課題】近年、膨大な情報が情報メディアに蓄積され送受信されてきている。利用者が情報を効率よく取得し処理するためには、利用者にとって有用だ（たとえば重要だ、急ぎだ、面白い）と思われる情報を選定することが重要である。しかし、より精度よく情報の属性（たとえば重要性や緊急性など）の判定を行う場合は、情報の内容に踏み込んだ理解をすることが必要である。

【0004】一般に、テキスト情報の属性を完全に把握するためには、テキスト情報を自然言語処理技術により

分析する必要がある。しかし、自然言語処理技術を用い、情報の属性を完全に自動で解析することは不可能であるか、できたとしても膨大な時間が必要であり実用的でない。

【0005】通常、利用者にとって意味のある情報の取捨選択のために必要な属性を得るためにテキスト情報の全体が必要ではなく、多くの場合、情報に含まれるいくつかの表現が必要な属性を表わしている。

【0006】そこで、本発明の目的とするところは、情報に含まれるある属性を表わす表現を抽出し、それぞれの表現に対して適当な手段により処理し、処理結果を情報の取捨選択に用いることにより、より精度良く高速に情報の属性を判定する情報の属性判定システム及び方法並びに情報の属性判定プログラムを記録した記録媒体を提供することにある。

【0007】

【課題を解決するための手段】上記目的を達成するために本発明の情報の属性判定システムは、テキストもしくはテキストに変換可能な要素から構成される情報に含まれる情報の属性を表わす言葉抽出手段と、この言葉抽出手段で抽出された言葉に所

定のアルゴリズムにしたがって数値を付ける第1の数値付与手段と、この第1の数値付与手段で得られた数値を他の所定のアルゴリズムにしたがって組み合わせ別の数値を得る第2の数値付与手段と、この第2の数値付与手段で得られた数値によって当該情報の属性を定量的に判定する判定手段とを具備することを特徴とするものである。

【0008】また本発明の情報の属性判定方法は、テキストもしくはテキストに変換可能な要素から構成される情報に含まれる情報の属性を表わす言葉を当該情報から抽出するテキストデータ参照ステップと、このテキストデータ参照ステップで抽出された言葉を表現テーブルと参照し、表現テーブルにある表現が存在するか否かを判別する表現テーブル参照ステップと、この表現テーブル参照ステップで判別された言葉を表現データベースと参照し、表現データベース値による値付けをする表現データベース参照ステップと、この表現データベース参照ステップで得られた数値を計算式データベースと参照し、表現ごとの値を計算式へ代入して当該情報の値の計算をする計算式参照ステップと、この計算式参照ステップで得られた数値によって当該情報の属性を定量的に判定する判定ステップとを具備することを特徴とする。

【0009】また本発明の情報の属性判定プログラムを記録した記録媒体は、コンピュータに、テキストもしくはテキストに変換可能な要素から構成される情報に含まれる情報の属性を表わす言葉を当該情報から抽出する第1の手順、この第1の手順で抽出された言葉に表現ごとの値付けをする第2の手順、この第2の手順で得られた表現ごとの値を計算式へ代入して当該情報の値の計算をする第3の手順、この第3の手順で得られた数値によって当該情報の属性を定量的に判定する第4の手順、を実行させるための情報の属性判定プログラムを記録したことを特徴とする。

【0010】

【発明の実施の形態】以下図面を参照して本発明の実施の形態例を詳細に説明する。図1及び図2は本発明の一実施形態例を示すフロー図である。本発明の情報の属性判定システムは、テキストもしくはテキストに変換可能な要素から構成される情報に対して、それに含まれる情報の属性を表わす言葉を当該情報から抽出して、その言葉に図1のアルゴリズムにしたがって数値を付けること、および得られた数値をさらに図2のアルゴリズムにしたがって組み合わせ別の数値を得ること、および最後に得られた数値によって当該情報の属性により情報の属性（たとえば重要性、緊急性）の値を計算して当該情報の属性を定量的に判定する。

【0011】すなわち、図1に示すように、先ず、テキストデータ参照ステップでは、テキストデータを参照し、テキストもしくはテキストに変換可能な要素から構成される情報に含まれる情報の属性を表わす言葉を当該

情報から抽出する。次に、表現テーブル参照ステップでは、前記テキストデータ参照ステップで抽出された言葉を表現テーブルと参照し、表現テーブルにある表現が存在するか否かを判別する。次に、表現テーブルにある表現が存在する場合には、表現データベース参照ステップにおいて、前記表現テーブル参照ステップで判別された言葉を表現データベースと参照し、表現データベース値による値付けをし、当該情報における表現ごとの値を得る。なお、表現テーブルにある表現が存在しない場合には、表現データベースの表現がない場合の既定値を参照して値付けをし、当該情報における表現ごとの値を得る。前記表現テーブルには、表現の種類、表現の形式、…が収容され、前記表現データベースには、「表現の種類、表現、値、…」及び「表現の種類、ない場合、値、…」という組みで登録される。

【0012】次に、図2に示すように、表現テーブル参照ステップでは、図1のアルゴリズムにしたがって得られた数値に対応する表現を抽出する。次に、計算式参照ステップでは、情報の対象領域における計算式とその係数データベースを参照し、当該情報における表現ごとの値を得るため、表現ごとの値を計算式へ代入し、当該情報の値の計算をして当該情報の属性を定量的に判定する当該情報の判定値を得る。

【0013】なお、コンピュータに、テキストもしくはテキストに変換可能な要素から構成される情報に含まれる情報の属性を表わす言葉を当該情報から抽出する第1の手順、この第1の手順で抽出された言葉に表現ごとの値付けをする第2の手順、この第2の手順で得られた表現ごとの値を計算式へ代入して当該情報の値の計算をする第3の手順、この第3の手順で得られた数値によって当該情報の属性を定量的に判定する第4の手順、を実行させるための情報の属性判定プログラムを記録した記録媒体を構成してもよい。

【0014】

【実施例】以下、本発明の実施例として、ユーザサポートの業務電子メールを対象にした情報の重要性、緊急性の計算例を示す。図3の例のように、利用者が「至急、明後日の電話工事のスケジュールを連絡してください。○○会社××」という電子メールを受信するとき、自動的にその電子メール1通ごとに、図8に示すような表現テーブルからある属性を表わす表現「受信者と発信者の関係」「キーワード」「期日」「緊迫表現」の記述を抽出する。その記述ごとに、それに対する値を図9～図12のそれぞれの表現データベースから検索し、計算式を適用する。図9はキーワードによる値の例であり、図10は緊迫表現による値の例であり、図11は受信者と発信者の関係への値の例であり、図12は期間の長さによる値の例である。

【0015】すなわち、受信者と発信者の関係は、「関連会社」で $\alpha = 3$ であり、キーワードは、「電話工事」

で $\beta = 5$ である。したがって、(興味度) = $a \times \alpha + b \times \beta = 8$ となる。なお、係数 a 、 b は1とする。

【0016】また、期日から到着日の期間の長さは、「明後日」で $\gamma = 1$ であり、緊迫表現は、「至急」で $\delta = 5$ である。したがって、(緊迫度) = $c \times \gamma + d \times \delta = 9$ となる。なお、係数 c 、 d は1とする。

【0017】以上より、(優先度)² = (興味度)² + (緊迫度)² = $8^2 + 9^2 = 64 + 81 = 145$

また、図4の例のように、利用者が「こんにちわ、○○商品のご紹介です。××のホームページを見てください。明日までです。」という電子メールを受信するとき、自動的にその電子メール1通ごとに、図8に示すような表現テーブルからある属性を表わす表現「受信者と発信者の関係」「キーワード」「期日」「緊迫表現」の記述を抽出する。その記述ごとに、それに対する値を図9～図12のそれぞれの表現データベースから検索し、計算式を適用する。

【0018】すなわち、受信者と発信者の関係は、なしで $\alpha = 2$ であり、キーワードは、なしで $\beta = 3$ である。したがって、(興味度) = $a \times \alpha + b \times \beta = 5$ となる。なお、係数 a 、 b は1とする。

【0019】また、期日から到着日の期間の長さは、なしで $\gamma = 1$ であり、緊迫表現は、「明日まで」で $\delta = 5$ である。したがって、(緊迫度) = $c \times \gamma + d \times \delta = 6$ となる。なお、係数 c 、 d は1とする。

【0020】以上より、(優先度)² = (興味度)² + (緊迫度)² = $5^2 + 6^2 = 25 + 36 = 61$

図7は優先度を求める計算式の例であり、情報の重要性、緊急性を算出する例は図7の計算式である。ここで計算式中の興味度とは、その情報への利用者の興味の度合いのことであり、受信者と発信者の関係と、利用者の興味を表わすキーワードにより算出される。また緊迫度とは、利用者にとってその情報の緊迫の度合いのことであり、期日と緊迫表現により算出される。この興味度と緊迫度より、その情報の重要性、緊急性を算出したものが優先度である。また、「受信者と発信者の関係」「キーワード」「期日」「緊迫表現」のそれぞれ項目に合った記述が複数あった場合には、値の大きいを選択する。この「受信者と発信者の関係」「キーワード」による値から、図7の計算式の興味度を算出する。また、「期日」「緊迫表現」による値から、図7の計算式の緊迫度を算出する。この興味度と緊迫度により、優先度を算出する。これが1通の電子メールのもつ優先度となる。

【0021】この例では、図3では「受信者と発信者の関係」「キーワード」「期日」「緊迫表現」のすべての項目において、高い値が設定され計算されるが、図4では「緊迫表現」のみ高い値が設定され計算される。よって、利用者にとっての重要性、緊急性の順位は図3の電子メール、図4の電子メールの順である。ここで、言葉

への値付けと計算式中の係数(a、b、c、d)は、この計算式を適用する領域によって定まるものである。

【0022】次に、この情報の重要性、緊急性の判定を用いた、電子メールのシステムとして図5及び図6が考えられる。図5は、電子メールのサーバ上に構築した情報配信システムを表わしている。サーバの機能として、到着した電子メールの重要性、緊急性を計算しておくことによって、携帯端末と携帯電話などという端末性能・通信性能の低い環境において、重要、緊急のメールを選択的に取り出すことができ、到着しているすべての電子メールを取り出す必要がない。

【0023】すなわち、先ず、複数の電子メールから順次電子メールを取り出し、取り出した電子メールについて、キーワードデータベースからキーワードデータを取り出し、キーワードデータによる値付けをする。また、前記取り出した電子メールについて、発信者の組織や人間関係データベースから組織や人間関係データを取り出し、発信者の組織や人間関係データによる値付けをする。また、前記取り出した電子メールについて、緊迫表現データベースから緊迫表現データを取り出し、緊迫表現データによる値付けをする。また、前記取り出した電子メールについて、期日表現データベースから期間データを取り出し、期間データによる値付けをする。前記キーワードデータによる値付けと、前記発信者の組織や人間関係データによる値付けから興味度を計算する。また、前記緊迫表現データによる値付けと、前記期間データによる値付けから緊迫度を計算する。前記興味度と緊迫度から優先度を計算して複数の電子メールに優先度を付ける。利用者から(××通)電子メールの配信要求があった場合、複数の電子メールの中から(××通だけ)優先度順に送信する。

【0024】図6は、利用者端末の電子メールツールの機能を表わしている。時に応じて電子メールの重要性、緊急性を計算すれば、多数のメールの中から、その時に必要な情報を選択することが可能になる。

【0025】すなわち、先ず、サーバから順次電子メールを取り出し、取り出した電子メールについて、キーワードデータベースからキーワードデータを取り出し、キーワードデータによる値付けをする。また、前記取り出した電子メールについて、発信者の組織や人間関係データベースから組織や人間関係データによる値付けをする。また、前記取り出した電子メールについて、緊迫表現データベースから緊迫表現データを取り出し、緊迫表現データによる値付けをする。また、前記取り出した電子メールについて、期日表現データベースから期間データを取り出し、期間データによる値付けをする。前記キーワードデータによる値付けと、前記発信者の組織や人間関係データによる値付けから興味度を計算する。また、前記緊迫表現データによる値付けと、前記期間データによる値付

けから緊迫度を計算する。前記興味度と緊迫度から優先度を計算して複数の電子メールに優先度を付け、複数の電子メールの中から優先度順に表示する。

【0026】なお、図5の電子メールを音声メッセージとすれば、ネットワークに蓄積された音声メッセージ（留守番電話など）システムを表わしている。重要な販売店からの連絡や緊急の問い合わせなどを選択して、聞くことが可能になる。電話では蓄積された音声メッセージを一覧する方法に乏しいので有用な機能と考えられる。

【0027】

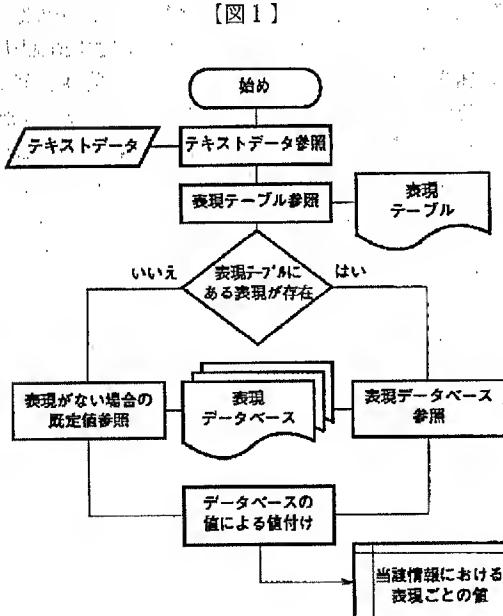
【発明の効果】以上説明したように本発明によれば、蓄積または送受信される情報に対してその情報の属性（たとえば重要性、緊急性）を高速に判定することが可能で、重要、緊急な情報のみを取得することや、必要な情報が不要な情報に埋もれてしまい見つけられなくなるということが回避でき、利用者が効率よく情報を取得し処理できるようになり有益である。

【図面の簡単な説明】

【図1】本発明の一実施形態例の一部を示すフロー図である。

【図2】本発明の一実施形態例の一部を示すフロー図である。

【図3】本発明の一実施例を示す説明図である。*



【図1】

* 【図4】本発明の他の実施例を示す説明図である。

【図5】本発明に係る電子メールのサーバ上に構築した情報配信システムを示すフロー図である。

【図6】本発明に係る利用者端末の電子メールツールの機能システムを示すフロー図である。

【図7】本発明に係る優先度を求める計算式の一例を示す説明図である。

【図8】本発明に係るある属性を表わす表現の一例を示す説明図である。

10 【図9】本発明に係るキーワードによる値の一例を示す説明図である。

【図10】本発明に係る緊迫表現による値の一例を示す説明図である。

【図11】本発明に係る受信者と発信者の関係への値の一例を示す説明図である。

【図12】本発明に係る期間の長さによる値の一例を示す説明図である。

【符号の説明】

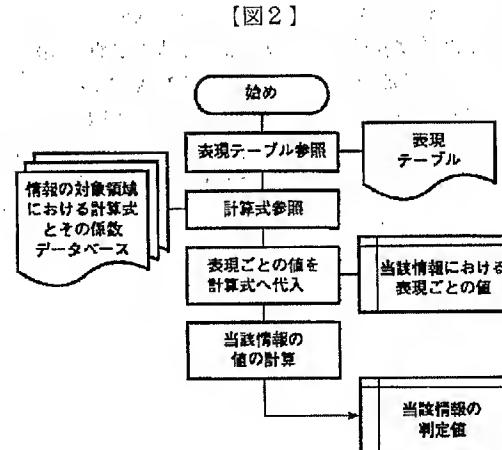
α 受信者と発信者の関係による値

β キーワードに対する値

γ 期日から到着日の期間の長さによる値

δ 緊迫表現に対する値

a, b, c, d 式を適用しようとしている対象領域における係数



【図2】

【図7】

$$\alpha = (\text{受信者と発信者の関係による値})$$

$$\beta = (\text{キーワードに対する値})$$

$$(\text{興味度}) = f_1(g(a, \alpha), h(b, \beta))$$

$$(\text{緊迫度}) = f_2((c, \gamma), (d, \delta))$$

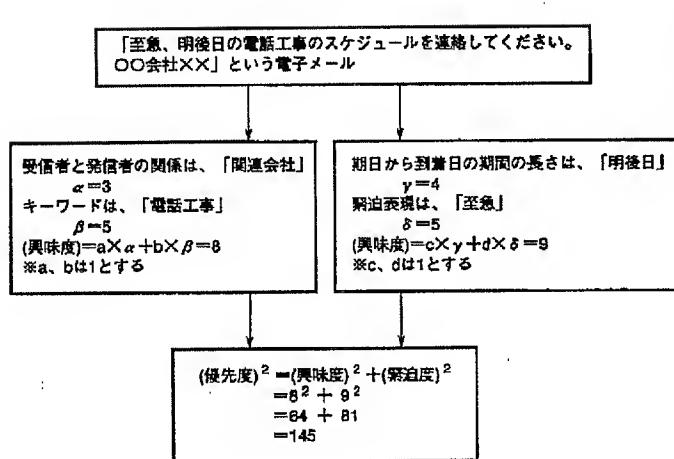
$$※ a, b, c, d は、式を適用しようとしている対象領域における係数$$

$$\gamma = (\text{期日から到着日の期間の長さによる値})$$

$$\delta = (\text{緊迫表現に対する値})$$

$$(\text{優先度}) = F((\text{興味度}), (\text{緊迫度}))$$

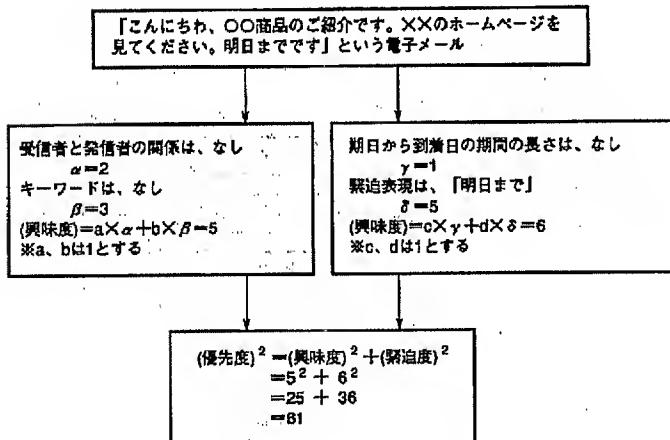
【図3】



【図9】

キーワード	値
仕事上よく用いる言葉(たとえば電話工事)など	5
職種上よく用いる言葉など	4
上記以外	3

【図4】

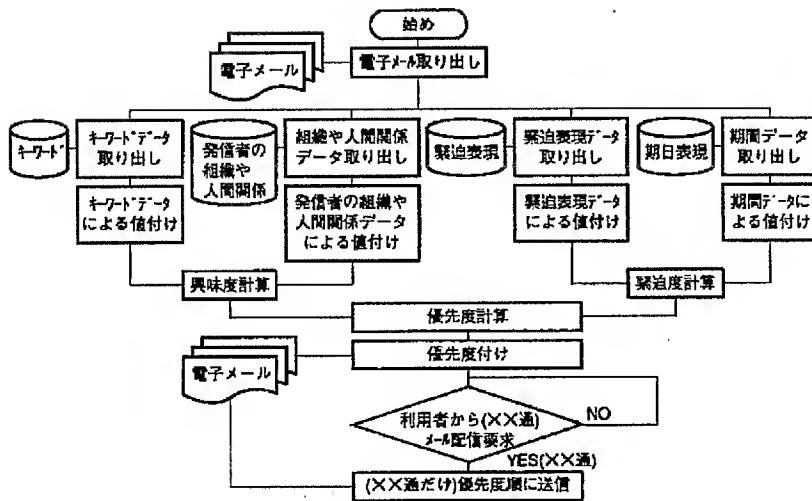


【図10】

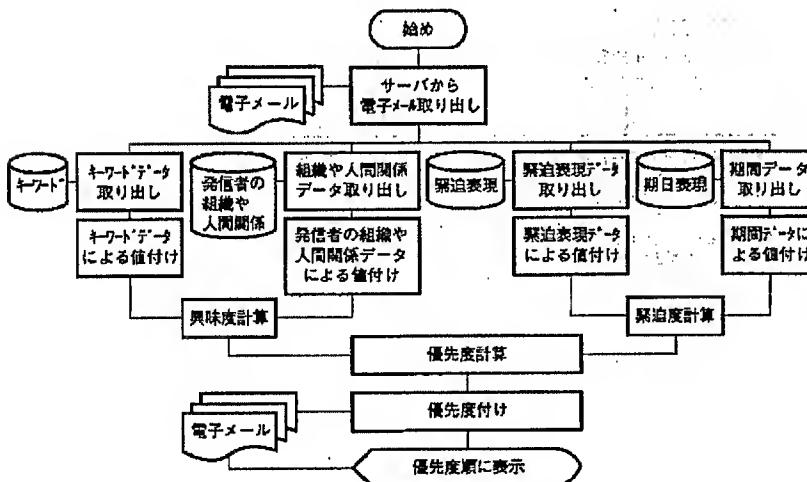
緊迫表現	値
問い合わせです、回答願います、注意して、出発する、返事を下さい、至急、急いで、など	5
お願ひいたします、教えて、伺います、報告して、困った、うまくいきません、質問です、など	4
お知らせください、ご返事ください、など	3
してください、など	2
上記以外	1

【図8】

【図5】



【図6】



【図11】

受信者と発信者の関係	値
同一フロア内、直属上司(??部 ××課長)など	5
同一会社(??部 ××課長)など	4
関連会社など	3
上記以外	2

【図12】

期日から到着日の期間の長さ	値
期日まで1日以内、期日から1日以内	5
期日まで3日以内	4
期日まで1週間以内	3
期日まで2週間以内	2
期日まで2週間以上	1